Background

Cancer is the second leading cause of death among Americans. One of every four deaths in the United States is due to cancer.[1,2] The American Cancer Society estimates that in 2005, about 1,372,910 Americans will receive a new diagnosis of invasive cancer, and 570,280 Americans will die of this disease. These estimates do not include *in situ* cancers or the more than 1 million cases of basal and squamous cell skin cancers expected to be diagnosed this year. The National Cancer Institute (NCI) recently estimated that on January 1, 2002, 10.1 million Americans were alive with a history of invasive cancer.[3]

According to the 2005 Annual Report to the Nation on the Status of Cancer, age-adjusted incidence rates for all cancers combined were stable from 1995 through 2002 in men, but rates in women have increased 0.3% annually since 1987. U.S. death rates for all cancer sites combined decreased beginning in 1993.[4] However, the number of Americans diagnosed with cancer each year is expected to double in the next 50 years, from 1.3 million to 2.6 million. The anticipated growth and aging of the U.S. population are factors that will increase the number of people who are diagnosed with and treated for cancer.[5]

The National Institutes of Health estimated that in 2005, the overall annual cost of cancer would be about $209.9 billion,[6] broken down as follows:

- Direct medical costs, including health expenditures: $74.0 billion.
- Indirect costs associated with lost productivity due to illness: $17.5 billion.
- Indirect costs associated with lost productivity due to premature death: $118.4 billion.

These costs are likely to increase because of the anticipated growth and aging of the U.S. population.

There are effective primary and secondary prevention measures that could substantially reduce the number of new cancer cases and prevent many cancer-related deaths. To reduce the nation's cancer burden, we must reduce behavioral and environmental factors that increase cancer risk, and we must ensure that high-quality screening services and evidence-based treatments are available and accessible, particularly to medically underserved populations.[7,8]

Cancer registries collect data about the occurrence of cancer (incidence), the types of cancer (morphology), the site in the body where the cancer first occurred (primary site), the extent of disease at the time of diagnosis (stage), the planned first course of treatment, and the outcome of treatment and clinical management (survival and vital status).[9,10] Cancer data are reported to metropolitan area, regional, and statewide cancer registries from a variety of medical facilities, including hospitals, physicians' offices, radiation facilities, freestanding surgical centers, and pathology laboratories. Cancer death data are recorded on death certificates that are sent to state vital statistics offices. Death certificates contain information regarding primary cancer site and morphology.

Information derived from population-based central cancer registries and from death certificates is critical for directing effective geographic area- or population-specific cancer prevention and control programs that focus on preventing behaviors that put people at increased risk for cancer (e.g., smoking) and on reducing environmental risk factors (e.g., occupational exposure to known carcinogens). This information is also essential for deciding which geographic areas should have cancer-screening programs and for making long-term plans for adequate diagnostic and treatment services. Pooled data at the national, regional, and state levels will help federal and state public health officials establish, prioritize, and monitor national public health surveillance initiatives and track progress toward the national goals and objectives set forth in *Healthy People 2010*,[11] which contains a set of health objectives for the nation for the first decade of the 21st century. For more information on *Healthy People 2010*, visit http://www.healthypeople.gov/document.

Federal Programs

Surveillance, Epidemiology, and End Results (SEER) Program

In 1971, Congress passed the National Cancer Act that mandated the collection, analysis, and dissemination of data useful for the prevention, diagnosis, and treatment of cancer.[12] This mandate led to the establishment of the Surveillance, Epidemiology, and End Results (SEER) Program.[13] For more than 30 years, the NCI's SEER Program has provided statistics regarding cancer incidence, survival, and mortality in the United States; monitored cancer incidence trends in geographic and demographic population groups; provided information on trends in extent of disease at diagnosis, therapy, and patient survival; promoted studies measuring progress in cancer control and etiology; provided specialty training in epidemiology, biostatistics, surveillance research, and tumor registry methodology, operations, and management; and developed new statistical methods, models, and software for the analysis and presentation of national and small-area statistics.

The SEER Program currently collects and publishes cancer incidence and survival data from 14 population-based cancer registries and 3 supplemental registries covering approximately 26% of the U.S. population (Appendix G). Information on more than 3 million *in situ* and invasive cancer cases is included in the SEER database, and approximately 170,000 new cases are added each year within SEER coverage areas. (See Appendix H for the first diagnosis year for which data were reported to NCI for each SEER area.) The mortality data reported by SEER are provided by the National Center for Health Statistics (NCHS). A public use data set is issued each year by the SEER Program for additional analyses.

For more information on the SEER Program, visit http://seer.cancer.gov.

National Program of Cancer Registries (NPCR)

Recognizing the need for more complete local, state, regional, and national cancer incidence data, Congress established the National Program of Cancer Registries (NPCR) in 1992 by enacting the Cancer Registries Amendment Act, Public Law 102-515; the program was reauthorized in 1998.[14] Congress mandated the Centers for Disease Control and Prevention (CDC) to provide funds to state and territorial health departments (or their authorized agencies) at a ratio of $3:$1 to match state support for the central cancer registry. As of 2005, CDC funds a total of 49 statewide and territorial cancer registries at two levels: capacity building for new registries or basic implementation for existing registries (Appendix G).

NPCR registries cover 96% of the U.S. population. NPCR has the state and national capacity to monitor the cancer burden; identify cancer incidence variation for racial and ethnic populations and for regions within a state, among states, and among regions; provide data for research; provide guidance for health resource allocation; respond to public concerns and inquiries about cancer; improve planning for future health care needs; and evaluate cancer prevention and control activities.[15]

In January 2001, NPCR registries began annually reporting their incidence data to CDC from the first diagnosis year for which a state or territorial cancer registry collected data with the assistance of NPCR funds (Appendix H). Data from the SEER metropolitan areas and SEER special population cancer registries operating in Alaska, Arizona, California, Georgia, Michigan, and Washington are reported to their respective NPCR state cancer registry for inclusion in the state's incidence data and are transmitted to CDC as part of the state's annual data submission. In January 2005, CDC received information on more than 9 million invasive cancer cases diagnosed during 1995–2002 and more than 1 million new invasive cancer cases are added each year. Two new data sets will be issued by NPCR for additional analyses in late 2005: (1) a public use data set of precalculated

cancer incidence rates will be available on CDC WONDER (http://wonder.cdc.gov), and (2) a restricted access file of individual record level data will be made available to selected state and national partners.

For more information on NPCR, visit http://www.cdc.gov/cancer/npcr.

National Vital Statistics System (NVSS)

The nation's vital statistics are available from the National Vital Statistics System (NVSS), which is maintained by NCHS. These vital statistics are provided through state-operated registration systems and are based on vital records filed in state vital statistics offices. The recording of vital events is the responsibility of the individual states and independent registration areas (e.g., District of Columbia, New York City, territories) in which the event occurs. Legal responsibility for the registration of vital events rests with the individual states. Through its Vital Statistics Cooperative Program, NCHS cooperates with state vital statistics offices to develop and recommend standard forms for data collection and model procedures to ensure uniform registration of the events monitored by the NVSS. Detailed annual data on births, deaths (including infant deaths), and fetal deaths are available for the United States and for states, counties, and other local areas. Data variables include cause of death, age, race, Hispanic origin, sex, marital status, place of birth, residence of decedent, education level, and place of death. Monthly provisional data on vital statistics are available for the United States and each state. A public use data set is issued each year by NCHS for additional analyses.

For more information on NCHS and its NVSS, visit http://www.cdc.gov/nchs.

Collaborating Partner

North American Association of Central Cancer Registries (NAACCR)

Both federally funded registry programs (i.e., the SEER Program and NPCR) work closely with the North American Association of Central Cancer Registries (NAACCR) to promote cancer incidence surveillance in the United States and Canada. Established in 1987, NAACCR is an organization of population-based cancer registries, governmental agencies, professional associations, and private groups in North America dedicated to its mission to develop and promote cancer registration standards; provide education and training; certify population-based cancer registries; evaluate and publish data; and promote the use of cancer surveillance data and systems for cancer control, epidemiologic research, public health programs, and patient care to reduce the burden of cancer in North America. All state and metropolitan area registries participating in NPCR and SEER, as well as all provincial and territorial registries in Canada, are members of NAACCR. A public online query system, CINA Plus Online, is updated annually with the most recent 5 years of incidence data (see http://www.naaccr.org/cinap). An incidence data file from 1995 forward is also updated annually for qualified researchers. For more information about this file, contact the NAACCR office (Appendix C).

In 1992, NAACCR began annual reviews of member registries' data for completeness, accuracy, and timeliness. In 1997, this process was formalized into a certification program, whereby registries report their data in December and NAACCR evaluates the data using standard, objective measures. Registries that meet the highest standards for data quality are recognized through certification.[16-18]

In 1997, when NAACCR evaluated 1995 incidence data, 9 NPCR registries and all 10 SEER registries were certified. Eight years later, when NAACCR evaluated the 2002 incidence in 2005, 34 NPCR registries, 4 NPCR/SEER registries, and 9 SEER registries were certified. (Data from San Francisco-

Oakland and San Jose-Monterey are combined and evaluated as the Greater Bay Area.)

For more information on NAACCR, visit http://www.naaccr.org.

Data Sources

Incidence Data

Data from the registries participating in NPCR were reported to CDC as of January 31, 2005. Data from registries in the SEER Program were reported to NCI as of November 1, 2004, and made available through the SEER Program public use data file released in April 2005 (http://www.seer.cancer.gov/publicdata). For this report, data from California, Kentucky, Louisiana, and New Jersey (states that are supported by both NPCR and SEER) are presented as reported to CDC as of January 31, 2005.

The primary source of cancer incidence data is medical records. Staff at health care facilities abstract cancer incidence data from patients' medical records, enter the data into the facility's own cancer registry if it has one, and then send the data to the regional or state registry. Both NPCR and SEER registries collect data using uniform data items and codes as documented by NAACCR. This uniformity ensures that data items collected by the two federal programs are comparable.[10,19] Information on primary site and histology was coded according to the *International Classification of Diseases for Oncology, Third Edition* (ICD–O–3)[20] and categorized according to the revised SEER recodes dated January 27, 2003, which define standard groupings of primary cancer sites (Appendix J, Table J.1).[3] NPCR and SEER cancer registries consider as reportable all incident cases with a behavior code of 2 (*in situ*, noninvasive) or 3 (invasive, primary site only) in the ICD–O–3 with the exception of *in situ* cancer of the cervix. Basal and squamous cell carcinomas of the skin are also excluded, with the exception of those on the skin of the genital organs.[20] To ensure as much comparability as possible between ICD–O–2 and ICD–O–3, the following exclusions were made: borderline tumors of the ovary, refractory anemias (shortage of blood cells that are unresponsive to treatment), and other myelodysplastic syndromes (conditions caused by abnormalities of the blood-forming cells in the bone marrow); pilocytic astrocytomas (slow-growing, cystic brain cancer) were included.[4] The January 2003 SEER recodes were used to ensure (1) consistent site-type definitions over time and (2) consistency with other published 2002 cancer incidence data.[3,21]

This report also contains data for two rare cancers, Kaposi sarcoma (KS) and mesothelioma. KS is a cancer of connective tissue such as cartilage, bone, fat, muscle, and blood vessels. Since the vast majority of KS cases have developed in association with human immunodeficiency virus (HIV) infection and acquired immunodeficiency syndrome (AIDS), HIV/AIDS is listed as the underlying cause of death. Therefore, KS death rates were not included in this report. Mesothelioma is a cancer that generally occurs in the chest, abdominal region, and areas surrounding the heart. It is typically associated with exposure to asbestos. Because KS and mesothelioma are considered separate cancers for this report, they were removed from counts of other primary cancer sites.

Childhood cancer incidence data are published in two formats. The first format is according to the SEER Modification of the *International Classification of Childhood Cancer* (ICCC) (Appendix J, Table J.2), consistent with previous United States Cancer Statistics reports. The ICCC, which is based on ICD–O–2, was published in 1996 by the International Agency for Research on Cancer (IARC). The ICCC presents childhood cancers in 12 groups classified primarily by morphology. The SEER modification, which affects the classification of nervous system and bone tumors (http://seer.cancer.gov/iccc), was chosen for compatibility with other published data on childhood cancer rates in the United States.

The second format is according to the SEER site recode, which is based primarily on cancer site; the incidence data are presented in this format to make them comparable with other published mortality data. This format allows the incidence data for childhood cancers to be categorized into the same groups as adult cancers. Although these groupings

are not as appropriate for children as they are for adults, they are necessary to allow comparisons between childhood incidence and childhood mortality.

In situ bladder cancers were recoded to invasive bladder cases because the information needed to distinguish between *in situ* and invasive bladder cases is not always available or reliable. Case counts and rates for invasive cancers are included in this report. This report also includes counts and rates for *in situ* breast cancer cases among women; these are reported separately and are not included in counts or rates for the "all sites" category. Case counts and rates for leukemias were calculated for acute and chronic lymphocytic leukemia, acute and chronic myeloid leukemia, and other leukemias, which include other myeloid/monocytic leukemias. Nonreportable cancers and cancers in patients of unknown sex or age were omitted from all calculations, but cases in patients of unknown race were included in the "all races" category. Counts of cases used in this report are listed in Appendices D and E.

For the first time, this year's report features incidence data on primary intracranial and central nervous system (CNS) tumors using site and histology groupings agreed upon by a consensus conference at the Society for Neuro-Oncology annual meeting in 2000.[22] Conference attendees agreed with the site definition proposed by the National Coordinating Council on Cancer Surveillance's (NCCCS) Brain Tumor Working Group formed in response to a request to investigate national, population-based incidence data on CNS tumors. The Benign Brain Tumor Cancer Registries Amendment Act changed NPCR's definition of reportable tumors to include benign CNS tumors. Subsequently, in addition to NPCR, both SEER and the Commission on Cancer (CoC) agreed to require reporting of nonmalignant brain tumors, beginning with cases diagnosed on or after January 1, 2004. Beginning this year, Table 1.3.1.1 will include malignant incident cases only. In 2007, cases for nonmalignant brain tumors will be added, and an increase in rates will be seen in the following histology groups and subgroups: (groups) tumors of the cranial and spinal nerves; tumors of the sellar region and (subgroups) unique astrocytoma variants; neuronal/glial, neuronal; meningioma; hemangioma.

Mortality Data

Cancer mortality statistics in this report are based on information from all death certificates filed in the 50 states and the District of Columbia and processed by NVSS at NCHS for deaths that occurred in 2002. The model for the state registration of deaths in 2002 is the 1989 revision of the *U.S. Standard Certificate of Death*.[23,24]

The cancer mortality data were compiled in accordance with World Health Organization (WHO) regulations, which specify that member nations classify and code causes of death in accordance with the current revision of the *International Classification of Diseases* (ICD). Effective with deaths that occurred in 1999, the United States began using the Tenth Revision of this classification (ICD–10).[25]

The ICD not only details disease classification but also provides definitions, tabulation lists, the format of the death certificate, and the rules for coding cause of death. Cause-of-death data presented in this report were coded by procedures outlined in annual issues of the NCHS Instruction Manuals.[26,27]

Tabulations of cause-of-death statistics are based solely on the underlying cause of death. The underlying cause is defined by WHO as "the disease or injury that initiated the train of events leading directly to death, or the circumstances of the accident or violence that produced the fatal injury."[25] The underlying cause of death is selected from the conditions entered by the physician in the cause-of-death section of the death certificate. Generally, more medical information is reported on death certificates than is directly reflected in the underlying cause of death. This information is captured in NCHS multiple cause-of-death statistics.[28-30]

Since 1968, NCHS has computerized the coding of the underlying cause of death in accordance with WHO rules. In this system, called "Automated Classification of Medical Entities" (ACME),[31] multiple cause-of-death codes serve as inputs to the computer software that selects the underlying cause of death. In addition, NCHS has developed two computer systems as inputs to ACME. Beginning with 1990 data, the Mortality Medical Indexing, Classification, and Retrieval (MICAR) system[32,33] has been applied to automate coding of multiple causes of death. Then, beginning with data year 1993, SuperMICAR, an enhancement of the MICAR system, was applied to allow for literal entry of the multiple cause-of-death text as reported by medical certifiers in the states. Records that cannot be automatically processed by MICAR or SuperMICAR are manually multiple-cause coded and then further processed through ACME. For the 2002 mortality statistics, approximately 77% of the nation's death records were multiple-cause coded using SuperMICAR, and 23% were processed using MICAR only.

For consistency with the cancer incidence data, cancer sites in mortality data were grouped according to the revised SEER recodes dated January 27, 2003.[3] Because NCHS uses different groupings for some sites, the death rates in this report may differ slightly from those published by NCHS. In addition, under the ICD, differences occur in mortality and incidence coding. For example, there are several codes for mesothelioma in ICD–10 (depending on the primary site). However, one code in ICD–O–3 captures all the primary sites that mesothelioma affects. SEER recodes for cancer mortality are listed in Appendix J. In addition, counts of cancer incidence cases are updated continuously, but cancer mortality data are not.

All states and the District of Columbia submitted part or all of their 2002 mortality data in electronic data files to NCHS. All states provided precoded cause-of-death data to NCHS except Illinois, Kentucky, Ohio, West Virginia, and the District of Columbia.[34] For 2002, all states submitted precoded demographic data (e.g., sex and race of the deceased) for all deaths. Mortality data for the entire United States refer to deaths that occurred within the United States; data for geographic areas are by the decedent's place of residence. Deaths among overseas Armed Forces personnel are not included.

One index of the quality of reporting causes of death is the proportion of death certificates coded to ICD–10 codes R00–R99 (i.e., symptoms, signs, and abnormal clinical and laboratory findings not elsewhere classified). Although deaths occur for which the underlying causes are impossible to determine, the proportion classified as R00–R99 indicates the care and consideration given to the cause-of-death statement by the medical certifier. This proportion also may be used as a rough measure of the specificity of the medical diagnoses made by the certifier in various areas. In 2002, the percentage of all reported deaths in the United States assigned to symptoms, signs, and abnormal clinical and laboratory findings not elsewhere classified, was 1.23%, which was lower than in 2000 (1.33%) and 2001 (1.34%), but higher than in 1999 (1.12%).[34] In general, from 1990 through 1999, the percentage of deaths from this cause for all ages combined was fairly stable (1.08%–1.18%). In addition, causes of death are more likely to be misclassified for populations other than white as symptoms, signs, and abnormal clinical and laboratory findings not elsewhere classified, and this misclassification may affect comparisons of cause-specific death data.[35]

Population Denominator Data

The population estimates for the denominators of incidence and death rates presented in this report are race-specific (all races, whites, blacks, Asians/Pacific Islanders, and American Indians/Alaska Natives), Hispanic-specific, and sex-specific county population estimates aggregated to the state or metropolitan-area level. The county population estimates that are incorporated into NCI's SEER*Stat software (http://www.seer.cancer.gov/seerstat) to calculate cancer incidence and death rates are available at http://www.seer.cancer.gov/popdata. The SEER*Stat population estimates are a slight modification of the annual time series of July 1 county population estimates (by age, sex, race, and Hispanic origin) produced by the Population Estimates Program of the U.S. Bureau of the Census (Census Bureau) with

support from NCI through an interagency agreement. The Census Bureau's population estimates and documentation of the procedures used to develop them are available at http://www.census.gov/popest/counties. The estimates used in this report are postcensal estimates for 2002 (based on the 2000 census) that include bridged single-race estimates derived from the multiple-race categories through collaboration between the Census Bureau and CDC's NCHS. For more information on the 2000 bridged population estimates, see http://www.cdc.gov/nchs/about/major/dvs/popbridge/popbridge.htm.

Documentation regarding modifications made by NCI to Census Bureau estimates is available at http://www.seer.cancer.gov/popdata. Briefly, the modification affects only population estimates for Hawaii and is based on information from the Epidemiology Program of the Hawaii Cancer Research Center. The modified population estimates were obtained by SEER using survey data collected by the Hawaii Department of Health to address concerns that the Native Hawaiian population had been undercounted in previous censuses. The "Hawaii-adjustment" to Census Bureau estimates has the net result of reducing the estimated white population and increasing the estimated Asian/Pacific Islander population in the state. Census Bureau estimates for the total population, black population, and American Indian and Alaska Native populations in Hawaii are unaffected.

Registry Eligibility Criteria

Cancer incidence data included in this report are from statewide or metropolitan area cancer registries that have high-quality cancer incidence data for 2002 as demonstrated by meeting the following data quality criteria for all cancer sites combined:

- *Case ascertainment is 90% or more complete.* The registry data include at least 90% of the expected, unduplicated cases where the expected cases are estimated by using methods developed by NAACCR.[17,18,21,36] Because some cancer patients receive diagnostic or treatment services at more than one reporting facility, cancer registries perform a procedure known as "unduplication" to ensure that each cancer case is counted only once.[37]
- *No more than 5% of cases are ascertained solely on the basis of a death certificate.* The proportion of cases ascertained solely on the basis of a death certificate, with no other information on the case available after the registry has completed a routine procedure known as "death clearance and followback,"[37-39] is another measure of the completeness of case ascertainment.
- *No more than 3% of cases are missing information on sex.*
- *No more than 3% of cases are missing information on age.*
- *No more than 5% of cases are missing information on race.*
- *At least 97% of the registry's records passed a set of single-field and interfield computerized edits.* Computerized edits are computer programs that test the validity and logic of data components. For example, if (a) a patient received a diagnosis of cancer in 1999, (b) the patient's age was reported as 80 years, and (c) the patient's year of birth was reported as 1942, a computerized edit could, without human intervention, identify these components as incompatible. The computerized edits applied to the data in this report were designed by the SEER Program for use by SEER registries. During the 1990s, these edits were expanded and incorporated into NAACCR standards (http://www.naaccr.org) and into the EDITS software designed and maintained by CDC (http://www.cdc.gov/cancer/edits/edits.htm).

Appendix I, Table I.1, shows state- and metropolitan area-specific measures for the data quality criteria used to determine which registries' data were eligible for inclusion in this report.

Populations Covered by This Report

Incidence data on more than 1.1 million invasive cancer cases (including approximately 12,000 cases among children younger than 20 years) diagnosed during 2002 and reported by 44 state cancer registries (35 NPCR, 4 NPCR/SEER, and 5 SEER), the District of Columbia (NPCR), and 6 SEER metropolitan areas (Atlanta, Detroit, Los Angeles, San Francisco-Oakland, San Jose-Monterey, and Seattle-Puget Sound) are included in this report. In total, the NPCR and SEER cancer registries whose data are included in this report cover 93% of the U.S. population (Figure 1) (93% of the white population, 89% of the black population, 96% of the Asian/Pacific Islander population, 95% of the American Indian/Alaska Native population, and 98% of the Hispanic population). Population coverage may be affected by the suppression of state incidence data if there are only 16 or fewer cases or if the state requests that the data be suppressed. (For more information, see the discussions under "Suppression of Rates and Counts at the State, Regional, Division, and National Levels," "Hispanics," and "American Indians/Alaska Natives.")

Mortality data on 557,271 deaths in 2002 from malignant neoplasms (i.e., cancers) as recorded in NVSS from the 50 states and the District of Columbia are included in this report; therefore, with regard to mortality data, 100% of the U.S. population is covered.[34] In 2002, malignant neoplasms caused 23% of the total deaths and were overall the second leading cause of death in the United States.[34] In Alaska, Maine, Minnesota, and Oregon, 2002 age-adjusted death rates indicated that cancer was the leading cause of death; age-adjusted death rates in 2002 were similar for cancers and diseases of the heart for Massachusetts, Montana, and Washington State.[34]

Statistical Methods

Crude and Age-Specific Incidence and Death Rates

Crude rates are available at http://www.cdc.gov/cancer/npcr/uscs. Age-specific rates are presented in this report and on the Web.

The crude and age-specific incidence rates equal the total number of new cancer cases diagnosed in 2002 in the population category of interest, divided by the at-risk population for that category, and multiplied by 100,000 (cancers by primary site) or by 1 million (ICCC groupings of childhood cancers).

The crude and age-specific death rates equal the total number of cancer deaths in 2002 in the population category of interest, divided by the at-risk population for that category, and multiplied by 100,000.

Age-Adjusted Incidence and Death Rates

The occurrence of many cancers increases with age, as does cancer mortality. The age distribution of a population (i.e., the number of people in particular age categories) can change over time and can be different in different geographic areas. Age-adjusting the rates ensures that differences in incidence or deaths from one year to another or from one geographic area to another are not due to differences in the age distribution of the populations being compared.

The standard population used to age-adjust the rates in this report is the 2000 U.S. standard population, in accordance with a 1998 U.S. Department of Health and Human Services recommendation.[40,41] The 2000 U.S. standard population is based on the proportion of the 2000 population in specific age groups (younger than 1 year, 1–4 years, 5–9 years, 10–14 years, 15–19 years, . . . 85 years or older). The proportions of the 2000 population in these age groups serve as weights for calculating age-adjusted incidence or death rates. However, NCHS uses a different set of age groups in its age-adjustment of death rates.

Therefore the cancer death rates in this report may differ slightly from those published by NCHS. In addition, the 2000 U.S. standard population weights are not race or sex specific, so they do not adjust for differences in race or sex between geographic areas or among populations being compared. They do, however, provide the basis for adjusting for differences in the age distributions across groups defined by sex, race, geography, or other categories.

The 2000 U.S. standard population weights used for this report are based on single years of age from the Census P25-1130 series estimates of the 2000 U.S. population. For the 5-year age groups, the single years of age populations are summed to form the 5-year age groups. For more information, visit http://seer.cancer.gov/stdpopulations/single_age.html. These standard weights are used to compute age-adjusted incidence and death rates by the method of direct standardization as implemented in NCI SEER*Stat software (http://www.seer.cancer.gov/seerstat) and are described as follows:[42]

If N_j is the number of incident cases diagnosed in 2002, or the number of cancer deaths in 2002, in age category j, and P_j is the population size at risk in age category j, then the incidence or death rate R_j in age category j is defined as

$$R_j = N_j/P_j$$

If w_j is the 2000 U.S. standard population weight for age category j, then the age-adjusted (directly standardized) incidence or death rate R_{adj} is computed as

$$R_{adj} = \Sigma_j\, w_j\, R_j \times 100{,}000$$

Note from the multiplier in the above formula that incidence or death rates are expressed as cases or deaths per 100,000 persons. For childhood cancers coded according to ICCC, the multiplier in the formula is 1,000,000 because the childhood cancer rates are expressed per million persons.

Confidence Intervals

Confidence intervals reflect the range of variation in the estimation of the cancer rates. The width of a confidence interval depends on the amount of variability in the data. Sources of variability include the underlying occurrence of cancer as well as uncertainty about when cancer is detected and diagnosed, when a death from cancer occurs, and when the data about the cancer are sent to the registry or the state health department. In any given year, when large numbers of a particular cancer are diagnosed or when large numbers of cancer patients die, the effects of random variability are small compared with the large numbers, and the confidence interval will be narrow. With rare cancers, however, the rates are small and the chance occurrence of more or fewer cases or deaths in a given year can markedly affect the rate. Under these circumstances, the confidence interval will be wide to indicate uncertainty or instability in the cancer rate.

To estimate the extent of this uncertainty, a statistical framework is applied.[43] The standard model used for rates for vital statistics is the Poisson process,[44] which assigns more uncertainty to rare events relative to the size of the rate than it does to common events. The population risk profile is positioned to influence the underlying Poisson process from which rates arise, and only a single realization of that process is observed.

Parameters are estimated for the underlying disease process. For this report, we estimated a single parameter to represent the incidence rate and its variability. Of note, the Poisson model is capable of estimating separate parameters that represent contributions to the rate from various population risk factors, the effects of cancer control interventions, and other attributes of the population risk profile in any particular year.

For this report, we used confidence intervals that are expected to include the true underlying rate 95% of the time. Confidence intervals in this report were computed by using SEER*Stat Version 6.1.4 software (http://www.seer.cancer.gov/seerstat) and are based on the gamma method.[45] These gamma intervals are approximations to exact Poisson confidence intervals. The gamma method

performs better than other methods, especially for small incidence rates or for populations with age distributions that differ from the standard age distribution. Various factors such as population heterogeneity can sometimes lead to "extra-Poisson" variation in which the rates are more variable than would be predicted by a Poisson model. No attempt was made to correct for this. In addition, as discussed in "Interpreting the Data," the confidence intervals do not account for systematic (i.e., nonrandom) biases in the incidence rates.

Users of this report who want to know if the differences between the rates of various groups are statistically significant might check whether the confidence intervals overlap. However, we discourage the use of overlapping confidence intervals to test for statistically significant differences between two rates because the practice more frequently fails to detect significant differences than does standard hypothesis testing.[46]

Another consideration when comparing differences between rates is their public health importance. For some rates in this report, numerators and denominators are large and standard errors are therefore small, resulting in statistically significant differences that may be so small as to lack importance for decisions related to population-based public health programs.

Suppression of Rates and Counts at the State, Regional, Division, and National Levels

When the numbers of cases or deaths used to compute rates are small, those rates tend to have poor reliability.[44] Therefore, to discourage misinterpretation or misuse of rates or counts that are unstable because case or death counts are small, incidence and death rates and counts are not shown in tables and figures if the case or death counts are less than 16. A count of less than about 16 results in a standard error of the rate that is approximately 25% or more as large as the rate itself. Similarly, a case count of less than approximately 16 results in the width of the 95% confidence interval around the rate being at least as large as the rate itself. These relationships were derived under the assumption of a Poisson process and with the standard population age distribution assumed to be similar to the observed population age distribution.

Another important reason for using a cell suppression threshold value is to protect the confidentiality of patients whose data are included in a report by reducing or eliminating the risk of identity disclosure.[47,48] The cell suppression threshold value of 16, which was selected to reduce misuse and misinterpretation of unstable rates and counts in this report, is more than sufficient to protect patient confidentiality given the low level of geographic and clinical detail provided in the report.[49]

Because the incidence and death rates shown in the state-, sex-, and race-specific bar graphs in Figures 3.1.M1 through 3.58.F2 are presented in rank order, we applied a data suppression criterion in addition to the threshold value of 16 cases. In these figures, incidence rates are not ranked or shown for any sex-specific population groups of less than 50,000 people.

U.S. Census Regions and Divisions

Rates for U.S. Census regions and divisions were calculated by aggregating data reported from the states in each region and division. Only data from state registries that met the criteria for inclusion in this report (see "Registry Eligibility Criteria") were included in calculations of incidence rates for U.S. Census regions and divisions. There is a potential for bias in the incidence rates for Census regions and divisions where data for some states were excluded. We estimated cancer rates for Census regions or divisions with ineligible cancer registries by assuming that the incidence-to-mortality ratio in the portion of the region or division that is covered by eligible registries is the same as the incidence-to-mortality ratio in the portion that is not covered by eligible cancer registries. The age-adjusted incidence rates for U.S. Census regions and divisions are reported only if (1) at least 80% of the population for the Census region or division is covered by cancer registries that meet the criteria for inclusion in this report and (2) the 95% confidence intervals around the observed age-adjusted region or division incidence rates based on data from eligible

registries for each of six major cancer sites (prostate, female breast, male colorectal, female colorectal, male lung and bronchus, female lung and bronchus) included the estimate of the region or division rate calculated using the methods described in Appendix L.

On the basis of these analyses, we present in this report the observed age-adjusted incidence rates for all U.S. Census regions and divisions with the exception of the South and its divisions (South Atlantic, East South Central, and West South Central). The estimate of rates based on the methods described in Appendix L are not presented in this report but are used for determining the exclusion of observed age-adjusted rates for Census regions and divisions.

Case counts for U.S. Census regions and divisions are in Appendix E if all state cancer registries in the region or division met the criteria for inclusion in this report unless the count for exactly one state in the region or division is suppressed due to a count of less than 16.

Total United States

Cancer incidence rates for the United States are aggregate rates based on more than 1.2 million cancer cases reported from central cancer registries in 44 states, 6 metropolitan areas, and the District of Columbia. The United States rate calculated using the methods described in Appendix L was higher than the upper confidence limit for the observed rate for both male prostate and male lung and bronchus cancers. The same statistical criteria that were applied to rates and counts for U.S. Census regions and divisions were also applied to the rates for the entire United States (see "U.S. Census Regions and Divisions" and Appendix L). Although the cancer rates for the entire United States do not meet these criteria, we report the observed cancer rates for the 93% of the U.S. population covered by eligible cancer registries. As of 2005, these rates are the best estimates of the U.S. cancer burden available that are based on observed data.

Case counts for the U.S. incidence rates for all ages combined are in Appendix D, Table D.1.1M and Table D.1.1F. The U.S. case counts are provided only to allow readers to verify the crude rates (available at http://www.cdc.gov/cancer/npcr/uscs) by recalculation. The U.S. counts and U.S. cancer incidence rates in this report pertain to the 93% of the U.S. population covered by eligible cancer registries.

Interpreting the Data

Age-adjusted and age-specific rates are presented in the print version of this report. Crude rates have also been calculated and can be found at http://www.cdc.gov/cancer/npcr/uscs. Crude rates are helpful in determining burden and specific needs for services for a given population, compared with another population, regardless of size. Crude rates are influenced by the underlying age distribution of the state's population. Even if two states have the same age-adjusted rates, the state with the relatively older population (as demonstrated by having a higher median age) will have higher crude rates because incidence or death rates for most cancers increase with increasing age. Ideally, crude, age-adjusted, and age-specific rates are all used to plan for population-based cancer prevention and control interventions.[41]

Incidence Data

Published age-adjusted cancer incidence rates for diagnosis years before 1999 were calculated by using the 1970 U.S. standard population; for mortality data, the 1940 standard population was used. Beginning with the publication of data for the 1999 diagnosis year, cancer incidence rates were age-adjusted to the 2000 U.S. standard population. This change conforms to U.S. Department of Health and Human Services policy for reporting death and disease rates.[40,41] This policy was motivated by a need to standardize age-adjustment procedures across government agencies.[40] The change to the 2000 U.S. standard updated the calculation of age-adjusted rates to more closely reflect the current age distribution of the U.S. population and the current burden of cancer. Because of the aging of the U.S. population, the 2000 U.S. standard population gives more weight to older age categories than did the 1940 and 1970 standard populations.[5,41]

Because cancer incidence increases with age, the change to the 2000 U.S. standard population resulted in higher incidence rates for most cancers. The data published here should not be compared with cancer incidence rates adjusted to different standard populations.

Incidence rates are also influenced by the choice of population denominators used in calculating the rates. Because some state health departments use customized state population projections when calculating incidence rates, the rates published in this report may differ slightly from those published by individual states.

The new population estimates based on the 2000 census improve the accuracy of cancer incidence rates (see "Population Denominator Data"). Previously reported overall cancer rates, when recalculated with the new denominators, do not appear to change appreciably (Dr. Francis P. Boscoe, New York State Cancer Registry, and Dr. Barry Miller, NCI, personal communication, July 2003). However, rates for geographic areas with small populations and for specific racial or ethnic populations appear to be affected to a greater degree and should be interpreted with caution.[50] Furthermore, since corrections to the population denominators extrapolated from the 1990 census were larger in the late 1990s than earlier in the decade, rates calculated for the late 1990s may be subject to more change than rates calculated for the early 1990s. We published *United States Cancer Statistics: 1999 Incidence* using extrapolated 1999 population estimates based on the 1990 census, with the expectation that the 1999 incidence rates would be revised when the intercensal (i.e., based on both the 1990 census and the 2000 census) race-specific population estimates for 1999 became available.[51] The incidence rates published in *United States Cancer Statistics: 1999 Incidence* have been revised to incorporate the modified 1999 population denominators and are available at http://www.cdc.gov/cancer/npcr/uscs.

Statistical bias can arise if, within a region, division, or country, the sub-area for which data are available has rates that are substantially different from the rates in the sub-area for which data are not available. Because of bias, rates for a U.S. Census region or division, or the country, may not meet statistical criteria for inclusion in this report. It is possible to have some statistical bias even if the percentage of coverage is high and large numbers of cases are recorded. Where coverage is less than 100%, merely increasing the percentage of the population covered may not reduce statistical bias unless the covered population is similar to the uncovered population in terms of cancer rates or proportions. The U.S. counts and rates in this report pertain to the 93% of the U.S. population covered by eligible cancer registries. The 7% of the population that is not covered by eligible cancer registries may have different cancer rates than does the 93% that is covered, so reported observed cancer rates may not be representative of the entire United States. Of note, however, rates for the entire United States that were estimated using the method described in Appendix L for the six specific major cancer sites (prostate, female breast, male colorectal, female colorectal, male lung and bronchus, and female lung and bronchus) and the rates for the same sites that are published in this report did not differ by more than 1.8%. This observation provides strong support that the reported rates are representative of those for the entire United States.

Data quality is routinely evaluated by NPCR and the SEER Program.[52,53] Some evaluation activities are conducted intermittently to find missing cases or to identify errors in the data. Although the cancer registries whose data are included in this report meet data quality criteria for all invasive sites combined, the completeness and quality of site-specific data may vary. The observed rates may have been influenced by differences in the timeliness, completeness, and accuracy of the data from one registry to another, from one reporting period to another, and from one primary cancer site to another.

Completeness and accuracy of the site-specific data may also be affected by the time interval allowed for reporting data to the two federal programs. For this report of 2002 data, the NPCR and SEER time interval for reporting data differed by 3 months. NPCR allowed an interval of 25 months after the close of the diagnosis year (data submission by January 31, 2005), and SEER

allowed a shorter interval of 22 months after the close of the diagnosis year (data submission by November 1, 2004).

Delays in reporting cancer cases can affect timely and accurate calculation of cancer incidence rates.[52] Cases are reported continuously to state and metropolitan area cancer registries in accordance with statutory and contractual reporting requirements. After the initial submission of the most recent year's data to the federal funding agency, cancer registries continue to revise and update their data on the basis of new information received. Therefore, some cancer cases for the 2002 diagnosis year will likely have been reported to state and metropolitan area cancer registries after these registries submitted their 2002 data to CDC or NCI. For this reason, incidence rates and case counts reported directly by state or metropolitan area cancer registries may differ from those in this publication. Reporting delays appear to be more common for cancers that are usually diagnosed and treated in nonhospital settings such as physician offices (e.g., early-stage prostate and breast cancer, melanoma of the skin). NCI routinely models SEER reporting patterns and estimates that the delay-adjusted 2002 incidence rate for all sites combined is about 4% higher than the observed 2002 age-adjusted incidence rate. Delay adjustments for 2002 SEER age-adjusted rates vary: melanoma is 12%, prostate cancer is 5%, and breast cancer is 2% (Dr. Brenda K. Edwards, NCI, personal communication, August 2005). Updates to observed data and reported cancer rates are due to improvements in the registry database gained through additional knowledge that only comes with increased time and effort (Dr. Brenda K. Edwards, NCI, personal communication, August 2005). Methods to adjust incidence rates for reporting delay were not applied to the data in this report.[54]

Each year, not only do state cancer registries submit data for a new diagnosis year to CDC or NCI, but they also submit an updated version of previous years' data. Federal agencies in turn update their cancer incidence statistics with each data submission and document the states' data submission date whenever the data are published. These continual updates by state and federal agencies illustrate the dynamic nature of cancer surveillance and the attention to detail that is characteristic of cancer registries. Each year when *United States Cancer Statistics* is published, we publish updates to previous years' data on the Web. Users of cancer incidence data published by federal agencies should be mindful of the data submission date for all data used in their comparisons. See "HTML Edition of *United States Cancer Statistics*" for more information.

Geographic variation in cancer incidence rates may be the result of regional differences in the exposure of the population to known or unknown risk factors.[55-57] A comparison of incidence rates based on data from all NPCR registries combined and from all SEER registries combined revealed that the two federal programs had different cancer incidence rates for some cancer sites.[58] Higher incidence rates for tobacco-related cancers and for colon cancer were observed in all NPCR areas combined when compared with SEER areas combined. The relationship was reversed (i.e., rates were higher in all SEER areas combined) for melanoma, prostate cancer, female breast cancer, and cancers that are more common among Asians/Pacific Islanders (e.g., stomach cancer, cancer of the liver/intrahepatic bile duct). Rates for cancer sites classified as ill-defined were higher in all NPCR areas combined. These registry program differences may arise because of differences in sociodemographic population characteristics (e.g., age, race and ethnicity, geographic region, urban or rural residence), screening use, health-related behaviors (e.g., behaviors related to tobacco use, diet, physical activity), exposure to cancer-causing agents, or registry operations factors (e.g., completeness, timeliness, specificity in coding cancer sites). Cancer researchers are investigating variability associated with known factors that affect cancer rates and risks by using model-based statistical techniques and other approaches for surveillance research. Differences in registry operations are being evaluated to ensure the consistency and quality in reporting data.

Mortality Data

The cancer mortality statistics in this report are influenced by the accuracy of information on the death certificate. Cause of death determined by autopsy combined with clinical data is considered

the best estimate of the true cause of death.[59] Autopsy studies of mortality data coded according to the eighth or ninth revision of ICD (ICD–8A or ICD–9) indicate that, when neoplasms (i.e., cancers) are an underlying cause of death, the sensitivity of death certificates was 87%–93%, and their predictive value positive was 85%–96%.[59-61] However, these studies are limited by selection bias, and currently less than 10% of deaths in the United States are autopsied.[62] The percentage of cancers coded as the underlying cause of death on the death certificate that agree with the cancer diagnosis in the medical record is an indication of the reliability with which underlying cause of death can be determined from the death certificate. Available studies show that 78%–85% of malignant neoplasms coded as an underlying cause of death on death certificates agreed with the clinical cancer diagnosis in medical records under ICD–8A or ICD–9,[63-65] with a range of 69% for larynx cancer to 98% for prostate cancer under ICD–9 (Appendix I, Table I.2). These results underscore the need to further monitor the accuracy of cancer mortality data overall and by anatomic site.

Some cancer patients may die with cancer (rather than die of it) as an underlying cause of death. Comparing the original cancer diagnosis in the medical record with those cancers later coded as an underlying cause of death on death certificates is a way of measuring if a person died with cancer rather than of it. Findings from an 11-year study under ICD–9 showed that about 83% of malignant neoplasms recorded on the medical record in ICD for oncology were also coded as an underlying cause of death from death certificates;[65] this percentage ranged from 72% for larynx cancer to 97% for multiple myeloma (Appendix I, Table I.2). The SEER study suggests that misattribution bias (i.e., the mistaken assignment of cancer as the underlying cause of death because the decedent received a diagnosis of cancer) affects how cancer is recorded on death certificates.[66]

In collaboration with the Social Security Administration and the National Association for Public Health Statistics and Information Systems, NCHS is developing a Model Vital Event Re-Engineered System to improve the accuracy and timeliness of vital statistics disseminated through the NVSS. Under the system, standard certificates for births and deaths will be revised, and state data systems will be re-engineered to better accommodate revisions, special studies or projects, and linkage with other health promotion programs. With regard to mortality statistics, handbooks have been revised for professionals who complete death certificates. In 2003, the standard certificate of death was revised.[67] For 2005, plans are in place for up to 21 states to adopt the revised death certificate.

Race and Ethnicity in Cancer Data

The NAACCR Race and Ethnicity Identifier Assessment Project confirmed the importance of publishing cancer rates by race and ethnicity.[68] In cancer incidence, race and ethnicity information is abstracted from medical records and then grouped into race and ethnicity categories.[19] Although state registries across the country use standardized data items and codes for both race and ethnicity (i.e., Hispanic origin), the initial collection of this information by health care facilities and practitioners and the procedures for assigning and verifying codes for race and ethnicity are not well standardized.[68] Thus, some inconsistency is expected in this information.

In cancer mortality, race and Hispanic origin are reported separately on the death certificate by the funeral director as provided by an informant or, in the absence of an informant, on the basis of observation.[34] Inconsistencies in the collection and coding of data on race and Hispanic origin and their effect on mortality statistics have been described previously.[69] The net effect of misclassification is an underestimation of deaths and death rates for races other than white or black. In addition, under-coverage of minority populations in the census and resultant population estimates introduce biases into death rates by race.[69-71] For the white population, published death rates are overstated by an estimated 1% and for the black population by 5%, resulting principally from undercounts of these populations in the census.

In this *United States Cancer Statistics* report, cancer incidence and mortality data are presented for all races combined and by race (whites, blacks, Asians/Pacific Islanders, and American

Indians/Alaska Natives) and ethnicity (Hispanics). Data for Asians/Pacific Islanders and American Indians/Alaska Natives are presented only for the nation and for states with at least 100,000 population because of concerns regarding possible misclassification of race data and the relatively small sizes of these populations in the United States (see Figures 3.1.M1–3.58.F2).

Asians/Pacific Islanders

For the first time, data for Asians/Pacific Islanders were included in *United States Cancer Statistics: 2000 Incidence*.[72] The Asian/Pacific Islander population in the United States is approximately 12.5 million or 4.4% of the 2002 U.S. population, substantially smaller than the white or black populations.[73] The Asian/Pacific Islander population is concentrated in several states: California, New York, Hawaii, Texas, New Jersey, Illinois, and Washington.[74,75]

Grouping Asians and Pacific Islanders into one racial population can mask differences in subpopulations. The U.S. Asian/Pacific Islander population is not a homogeneous group. Rather, it comprises many subpopulations that differ in language, culture, and length of residence in the United States.[73,76] The three largest Asian subpopulations in the United States are Chinese, Filipino, and Asian Indian.[75] Although state cancer registries have designated codes for race that allow them to document the occurrence of cancer in 23 different Asian/Pacific Islander subpopulations,[19] the subpopulations are grouped into a single Asian/Pacific Islander race group in this report because of small numbers and concerns regarding the possible misclassification of race for Asian/Pacific Islander subpopulations.

Studies show that a person self-reported as Asian/Pacific Islander in a census or survey was sometimes reported as white on the death certificate.[77,78] Death rates are understated for Asians/Pacific Islanders by approximately 11%.[69] Studies are under way to examine the misclassification of race for Asian/Pacific Islander subpopulations and the underreporting of Asian/Pacific Islander race in cancer incidence data (Dr. Holly L. Howe, NAACCR, personal communication, August 2005).

Hispanics

The Office of Management and Budget defines Hispanics (or Latinos) as persons of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race.[79] For the first time, data for Hispanics were included in *United States Cancer Statistics: 2001 Incidence and Mortality*.[80] Hispanics have one of the highest growth rates among minority groups in the United States with approximately 37.4 million in the 2002 U.S. population, similar in size to the U.S. black population.[81-83] The three largest Hispanic subpopulations living in the United States are Mexican, Puerto Rican, and Cuban. States with 1 million or more Hispanics include California, Texas, New York, Florida, Illinois, Arizona, and New Jersey.[81]

NAACCR convened an expert panel to develop a best practices approach to improving Hispanic identification and is addressing Hispanic origin misclassification in central cancer registries. NPCR registries assigned Hispanic ethnicity through the standardized use of the NAACCR Hispanic Identification Algorithm (NHIA).[84] NHIA uses a combination of NAACCR variables to directly or indirectly classify cancer cases as Hispanic for analytic purposes. Cases reported as having Spanish/Hispanic origin (as indicated by NAACCR data element 190 with values 1–6[19]) are directly identified as Hispanic in the dataset. Cases reported as non-Spanish/non-Hispanic, Spanish surname only, or unknown whether Spanish (NAACCR data element 190 with a value of 0, 7, or 9[19]) are evaluated for possible Hispanic ethnicity through indirect identification. The ultimate goal of the algorithm is to classify these cases as Hispanic or non-Hispanic on the basis of an evaluation of the strength of the birthplace, race, and/or surname associations with Hispanic ethnicity. After applying the NHIA, cases not classified as Hispanic are classified as non-Hispanic, leaving no cases with unknown Hispanic status. More detailed information on the algorithm is available at http://www.naaccr.org.[85]

In this report, NHIA-classified case counts and incidence rates for Hispanics are presented for 33 NPCR registries. The following states have opted

not to present state-specific, NHIA-classified Hispanic counts and rates: Florida, Kentucky, New Hampshire, Pennsylvania, South Carolina, and Wisconsin. Some registries may have a small percentage of Hispanics in their states. Quality reviews of NHIA and the data elements that make up the NHIA, including but not limited to Hispanic ethnicity, race, birthplace, surname, and maiden name for women, have been conducted.[86] National rates include data for registries that opted not to present state-specific, NHIA-classified Hispanic counts and incidence rates. Preliminary data analyses showed that exclusion of these registry data did not appreciably change the overall incidence rates.

In this report, counts and incidence rates for the 11 SEER registries are based on the NAACCR Spanish/Hispanic Origin data element only. Data for the Connecticut, Hawaii, Iowa, New Mexico, and Utah SEER registries are included in the national rate; counts and rates for Hawaii and metropolitan Detroit are not presented because of the small percentage of Hispanics in those areas.

Death counts and rates for Hispanics are presented at the national and state levels for all 50 states and for the District of Columbia. Hispanic origin is assigned to cancer mortality data on the basis of information collected from death certificates.

The National Longitudinal Mortality Study examined the reliability of Hispanic origin and reported a 89.7% record-by-record agreement and a net underreporting of Hispanic origin on death certificates by 7% compared with self-reports on the surveys.[69] Death rates for the Hispanic-origin population are also affected by under-coverage of this population group in the census and the resultant population estimates; the estimated net correction, taking into account both sources of bias, is 1.6%.[71]

American Indians/Alaska Natives

Over 560 American Indian tribes are recognized by individual states and the federal government.[8,87,88] The American Indian/Alaska Native population in the United States is approximately 2.8 million or 1.0% of the 2002 U.S. population, substantially smaller than the white or black populations and smaller than the Asian/Pacific Islander population.[89] The American Indian/Alaska Native population is concentrated in several states: California has 15% of the nation's American Indian/Alaska Native population; Arizona, 10%; Oklahoma, 10%; New Mexico, 7%; Texas, 5%; Alaska, 4%; New York, 4%; North Carolina, 4%; and Washington, 4%.[89]

Previous studies have found racial misclassification to contribute to lower death rates and lower cancer incidence rates among the American Indian/Alaska Native population. Based on a comparison of race reported on death certificates from 1979–1989 with nine Current Population Survey files for the years 1973–1985 from the National Longitudinal Mortality Study conducted by the U.S. Bureau of the Census, record-by-record agreement was only 57% for American Indians.[69] When the net agreement of counts by race was examined between the two sources, almost 40% more persons were reported as American Indian/Alaska Native in the Current Population Survey files than on the death certificates.[69] The range of underestimation of cancer incidence rates among this population are similar. Studies that estimate misclassification among American Indians/Alaska Natives using cancer registry data report these rates are underreported by 40%–57%, depending on the region of the country.[88,90,91]

Studies measuring racial misclassification with cancer registry data have linked cases with Indian Health Service (IHS) administrative records.[88,90,91] Similarly in 2004, NPCR cancer registries were required to link cancer cases diagnosed from 1995–2002 with the IHS patient registration database to identify American Indians/Alaska Natives that were classified in the registry as non-native. IHS provides medical services to American Indians/Alaska Natives who are members of federally recognized tribes, estimated to be approximately 55% of the American Indian/Alaska Native population (Dr. David Espey, IHS, personal communication, July 2005). IHS coverage of these populations varies by region, does not include American Indians/Alaska Natives who are members of non-federally recognized tribes, and underrepresents those who live in certain urban areas. However, American Indians/Alaska Natives who live outside of service counties may continue to receive IHS services or may have

received services before moving. Patient identifying information from cancer cases in state registries were linked to a database that included all American Indians/Alaska Natives that received medical care during 1985–2002 through IHS. Results of the linkage were captured in a new data element, IHS Link (NAACCR data element 192),[19] that was sent back to state cancer registries to include in their January 31, 2005, NPCR data submission.

In this report, national incidence counts and rates for American Indians/Alaska Natives for the 41 NPCR registries included in this report are based on Race1 (NAACCR data element 160)[19] and IHS Link. Counts and incidence rates for the five state SEER registries are based on Race1 only. California opted not to present state-specific American Indian/Alaska Native case counts, incidence rates, death counts, and death rates.

National death counts and rates consist of data obtained from all 50 states and the District of Columbia. American Indian/Alaska Native race for these data is obtained from information on the death certificate.

HTML Edition of *United States Cancer Statistics*

A printable version of *United States Cancer Statistics: 2002 Incidence and Mortality* is available on the Web at http://www.cdc.gov/cancer/npcr/uscs. The HTML edition replicates all the text, tables, and figures in this report and adds the following:

- Expanded Tables 1 and 2 with cancer incidence rates for men and women combined.
- Expanded Tables 1 and 2 to include crude rates.
- State rankings of incidence and mortality of selected major invasive cancers.

Previous versions of *United States Cancer Statistics* have been updated on the Web with the additional data mentioned previously as well as the following:

- Cancer mortality statistics for 1999 and 2000.
- 1999–2001 cancer cases as reported to CDC as of January 31, 2005, and as reported to NCI as of November 1, 2004, and made available through the SEER Program public use file (see "Interpreting the Data: Incidence Data"). The updated population coverage for incidence rates are as follows:
  - ➤1999 incidence: 93% of the U.S. population
  - ➤2000 incidence: 94% of the U.S. population
  - ➤2001 incidence: 91% of the U.S. population
- Cancer cases grouped according to the January 2003 modification to the SEER site recodes (see "Data Sources: Incidence Data").

All data presented in the HTML version can be downloaded in ASCII or Microsoft Excel format for use in other applications.

References

1. American Cancer Society. *Cancer Facts and Figures 2005.* Atlanta: American Cancer Society; 2005. Available at http://www.cancer.org.

2. Centers for Disease Control and Prevention. *The Burden of Chronic Diseases and Their Risk Factors: National and State Perspectives 2004.* Atlanta: Centers for Disease Control and Prevention; 2005. Available at http://www.cdc.gov/nccdphp/burdenbook2004/index.htm.

3. Ries LAG, Eisner MP, Kosary CL, Hankey BF, Miller BA, Clegg L, Mariotto A, Feuer EJ, Edwards BK. *SEER Cancer Statistics Review, 1975–2002.* Bethesda, MD: National Cancer Institute; 2005. Available at http://seer.cancer.gov/csr/1975_2002.

4. Edwards BK, Brown ML, Wingo PA, Howe HL, Ward E, Ries LAG, Schrag D, Jamison PM, Jemal A, Wu XC, Friedman C, Harlan L, Warren J, Anderson RN, Pickle LW. Annual report to the nation on the status of cancer, 1975–2002, featuring population-based trends in cancer treatment. *Journal of the National Cancer Institute.* 2005;97(19):1407-27.

5. Edwards BK, Howe HL, Ries LA, Thun MJ, Rosenberg HM, Yancik R, Wingo PA, Jemal A, Feigal EG. Annual report to the nation on the status of cancer, 1973–1999, featuring implications of age and aging on U.S. cancer burden. *Cancer* 2002;94(10):2766-92.

6. National Heart, Lung, and Blood Institute. *NHLBI Fact Book*, *Fiscal Year 2004*. Bethesda, MD: National Heart, Lung, and Blood Institute; 2005. Available at http://www.nhlbi.nih.gov/about/factpdf.htm.

7. Curry SJ, Byers T, Hewitt M. *Fulfilling the Potential of Cancer Prevention and Control.* Washington, DC: The National Academies Press; 2003.

8. Haynes MA, Smedley BD. *The Unequal Burden of Cancer: An Assessment of NIH Research and Programs for Ethnic Minorities and the Medically Underserved.* Washington, DC: The National Academies Press; 1999.

9. American College of Surgeons Commission on Cancer. *Standards of the Commission on Cancer*, *Volume II: Registry Operations and Data Standards (ROADS).* Chicago: American College of Surgeons; 1998.

10. Fritz A, Ries LAG. *The SEER Program Code Manual. 3rd ed.* Bethesda, MD: National Cancer Institute; 1998.

11. U.S. Department of Health and Human Services. *Healthy People 2010*, *2nd ed. 2 vols.* Washington, DC: U.S. Government Printing Office; 2000. Available at http://www.healthypeople.gov/document.

12. The National Cancer Act of 1971, Public Law No. 92-218, Stat. 1828, (December 23, 1971). Available at http://www3.cancer.gov/legis/1971canc.html.

13. Hankey BF, Ries LA, Edwards BK. The Surveillance, Epidemiology, and End Results Program: a national resource. *Cancer Epidemiology Biomarkers and Prevention* 1999;8(12):1117-21.

14. Cancer Registries Amendment Act, Public Law No. 102-515, Stat. 3312, (October 22, 1992). Available at http://www.cdc.gov/cancer/npcr/npcrpdfs/publaw.pdf.

15. Centers for Disease Control and Prevention. *Program Announcement #00027*, *National Program of Cancer Registries.* Atlanta: Centers for Disease Control and Prevention; 1999.

16. North American Association of Central Cancer Registries. *NAACCR Registry Certification.* Springfield, IL: North American Association of Central Cancer Registries; 2005. Available at http://www.naaccr.org/index.asp?Col_SectionKey=12&Col_ContentID=54.

17. Tucker TC, Howe HL, Weir HK. Certification of population-based cancer registries. *Journal of Registry Management* 1999;26(1):24-7.

18. Tucker TC, Howe HL. Measuring the quality of population-based cancer registries: the NAACCR perspective. *Journal of Registry Management* 2001;28(1):41-4.

19. Havener L, Hulstrom D. *Standards for Cancer Registries*, *Volume II: Data Standards and Data Dictionary*, *10th ed.*, *Version 11.* Springfield, IL: North American Association of Central Cancer Registries; 2004.

20. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin DM, Whelan S. *International Classification of Diseases for Oncology*, *Third Edition.* Geneva, Switzerland: World Health Organization; 2000.

21. Ellison JH, Wu X, Howe HL, McLaughlin C, Lake A, Firth R, Sullivan SK, Roney D, Cormier M, Leonfellner S, Kosary CL. *Cancer in North America*, *Volume One: Incidence.* Springfield, IL: North American Association of Central Cancer Registries; 2005. Available at http://www.naaccr.org/index.asp?Col_SectionKey=11&Col_ContentID=50.

22. McCarthy BJ, Surawicz T, Bruner JM, Kruchko C, Davis F. Consensus conference on brain tumor definition for registration. *Neuro-oncology* 2002;4(2):134-45.

23. National Center for Health Statistics. *Vital Statistics of the United States, 1989. Volume II, Mortality, Parts A and B.* Hyattsville, MD: National Center for Health Statistics; 1992. Available at http://www.cdc.gov/nchs/products/pubs/pubd/vsus/1985/1985.htm.

24. Tolson GC, Barnes JM, Gay GA, Kowaleski JL. The 1989 revision of the U.S. standard certificates and reports. *Vital and Health Statistics, Series 4* 1991;(28):1-34.

25. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems. 10th Revision.* Geneva, Switzerland: World Health Organization; 1992.

26. National Center for Health Statistics. *National Vital Statistics System Instruction Manual, Part 2a—Instructions for Classifying the Underlying Cause of Death, ICD–10, 2005.* Hyattsville, MD: National Center for Health Statistics; 2005. Available at http://www.cdc.gov/nchs/about/major/dvs/im.htm.

27. National Center for Health Statistics. *National Vital Statistics System Instruction Manual, Part 2b—Instructions for Classifying Multiple Causes of Death, ICD–10, 2005.* Hyattsville, MD: National Center for Health Statistics; 2005. Available at http://www.cdc.gov/nchs/about/major/dvs/im.htm.

28. National Center for Health Statistics. *Mortality Data, Multiple Cause-of-Death Public-Use Data Files.* Hyattsville, MD: National Center for Health Statistics; 2005. Available at http://www.cdc.gov/nchs/products/elec_prods/subject/mortmcd.htm#content.

29. Chamblee RF, Evans MC. TRANSAX: the NCHS system for producing multiple cause-of-death statistics, 1968–78. *Vital and Health Statistics, Series 1* 1986;(20):1-83.

30. Israel RA, Rosenberg HM, Curtin LR. Analytical potential for multiple cause-of-death data. *American Journal of Epidemiology* 1986;124(2):161-79.

31. National Center for Health Statistics. *National Vital Statistics System Instruction Manual, Part 2c—ICD–9 ACME Decision Tables for Classifying Underlying Cause of Death.* Hyattsville, MD: National Center for Health Statistics; 2001. Available at http://www.cdc.gov/nchs/about/major/dvs/im.htm.

32. National Center for Health Statistics. *National Vital Statistics System Instruction Manual, Part 2g—Data Entry Instructions for the Mortality Medical Indexing, Classification, and Retrieval System (MICAR), 1996–1997, ICD–9.* Hyattsville, MD: National Center for Health Statistics; 1996. Available at http://www.cdc.gov/nchs/about/major/dvs/im.htm.

33. National Center for Health Statistics. *National Vital Statistics System Instruction Manual, Part 2h—Dictionary of Valid Terms for MICAR.* Hyattsville, MD: National Center for Health Statistics; 1996.

34. Kochanek KD, Murphy SL, Anderson RN, Scott C. Deaths: final data for 2002. *National Vital Statistics Reports* 2004;53(5):1-116.

35. Becker TM, Wiggins CL, Key CR, Samet JM. Symptoms, signs, and ill-defined conditions: a leading cause of death among minorities. *American Journal of Epidemiology* 1990;131(4): 664-8.

36. Howe HL. *Conclusions of the Workgroup for High-Quality Criteria for Data Use: The NAACCR Narrative.* Springfield, IL: North American Association of Central Cancer Registries; 2001.

37. Havener L. *Standards for Cancer Registries, Volume III: Standards for Completeness, Quality, Analysis, and Management of Data.* Springfield, IL: North American Association of Central Cancer Registries; 2004.

38. Menck HR, Phillips JL. Central cancer registries. In: Hutchinson CL, Menck HR, Burch M, Gottschalk R, editors. *Cancer Registry Management: Principles and Practice. 2nd ed.* Dubuque, IA: Kendall/Hunt Publishing Company; 2004.

39. Seiffert JE, Hoyler SS, McKeen K, Potts M. Casefinding, abstracting, and death clearance. In: Menck HR, Smart C, editors. *Central Cancer Registries: Design, Management, and Use*. Chur, Switzerland: Harwood Academic Publishers; 1994.

40. Anderson RN, Rosenberg HM. Report of the second workshop on age adjustment. *Vital and Health Statistics, Series 4* 1998;(30):1-37.

41. Anderson RN, Rosenberg HM. Age standardization of death rates: implementation of the year 2000 standard. *National Vital Statistics Reports* 1998;47(3):1-16.

42. Breslow NE, Day NE. *Statistical Methods in Cancer Research, Volume II: The Design and Analysis of Cohort Studies*. Lyon, France: International Agency for Cancer Research; 1987.

43. Särndal C-E, Swennson B, Wretman J. *Model-Assisted Survey Sampling*. New York: Springer-Verlag; 1992.

44. Brillinger DR. The natural variability of vital rates and associated statistics. *Biometrics* 1986;42(4):693-734.

45. Fay MP, Feuer EJ. Confidence intervals for directly standardized rates: a method based on the gamma distribution. *Statistics in Medicine* 1997;16(7):791-801.

46. Schenker N, Gentleman JF. On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician* 2001;55(3):182-6.

47. Federal Committee on Statistical Methodology. *Report on Statistical Disclosure Limitations Methodology (Statistical Working Paper 22)*. Washington, DC: Office of Management and Budget; 1994. Available at http://www.fcsm.gov/working-papers/spwp22.html.

48. Doyle P, Lane JI, Theeuwes JM, Zayatz LM. *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*. Amsterdam: Elsevier Science; 2001.

49. National Center for Health Statistics. *NCHS Staff Manual on Confidentiality*. Hyattsville, MD: National Center for Health Statistics; 1997. Available at http://www.cdc.gov/nchs/data/misc/staffmanual2004.pdf.

50. National Cancer Institute. *NCI Fact Sheet: How Changes in U.S. Census Counts Affect NCI Cancer Rates*. Bethesda, MD: National Cancer Institute; 2003. Available at http://www.cancer.gov/newscenter/pressreleases/Census2000.

51. U.S. Cancer Statistics Working Group. *United States Cancer Statistics: 1999 Incidence*. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention *and* National Cancer Institute; 2002.

52. Fritz A. The SEER Program's commitment to data quality. *Journal of Registry Management* 2001;28(1):35-40.

53. Hutton MD, Simpson LD, Miller DS, Weir HK, McDavid K, Hall HI. Progress toward nationwide cancer surveillance: an evaluation of the National Program of Cancer Registries, 1994–1999. *Journal of Registry Management* 2001;28(3):113-20.

54. Clegg LX, Feuer EJ, Midthune DN, Fay MP, Hankey BF. Impact of reporting delay and reporting error on cancer incidence rates and trends. *Journal of the National Cancer Institute* 2002;94(20):1537-45.

55. Centers for Disease Control and Prevention. *Behavioral Risk Factor Surveillance System Operational and User's Guide. Version 3.0*. Atlanta: Centers for Disease Control and Prevention; 2005. Available at http://www.cdc.gov/brfss/pubs/index.htm#users.

56. Devesa SS, Grauman DJ, Blot WJ, Pennello GA, Hoover RN. *Atlas of Cancer Mortality in the United States, 1950–1994*. Bethesda, MD: National Cancer Institute; 1999. Available at http://www3.cancer.gov/atlasplus.

57. Howe HL, Keller JE, Lehnherr M. Relation between population density and cancer incidence, Illinois, 1986–1990. *American Journal of Epidemiology* 1993;138(1):29-36.

58. Wingo PA, Jamison PM, Hiatt RA, Weir HK, Gargiullo PM, Hutton M, Lee NC, Hall HI. Building the infrastructure for nationwide cancer surveillance and control—a comparison between the National Program of Cancer Registries (NPCR) and the Surveillance, Epidemiology, and End Results (SEER) Program (United States). *Cancer Causes and Control* 2003;14(2):175-93.

59. Kircher T, Nelson J, Burdo H. The autopsy as a measure of accuracy of the death certificate. *The New England Journal of Medicine* 1985; 313(20):1263-9.

60. Engel LW, Strauchen JA, Chiazze L Jr, Heid M. Accuracy of death certification in an autopsied population with specific attention to malignant neoplasms and vascular diseases. *American Journal of Epidemiology* 1980;111(1): 99-112.

61. Schottenfeld D, Eaton M, Sommers SC, Alonso DR, Wilkinson C. The autopsy as a measure of accuracy of the death certificate. *Bulletin of the New York Academy of Medicine* 1982;58(9):778-94.

62. Sinard JH. Factors affecting autopsy rates, autopsy request rates, and autopsy findings at a large academic medical center. *Experimental and Molecular Pathology* 2001;70(3):333-43.

63. Gittelsohn A, Senning J. Studies on the reliability of vital and health records: I. Comparison of cause of death and hospital record diagnoses. *American Journal of Public Health* 1979;69(7):680-9.

64. Percy C, Stanek E III, Gloeckler L. Accuracy of cancer death certificates and its effect on cancer mortality statistics. *American Journal of Public Health* 1981;71(3):242-50.

65. Percy CL, Miller BA, Gloeckler L, Ries LA. Effect of changes in cancer classification and the accuracy of cancer death certificates on trends in cancer mortality. *Annals of the New York Academy of Sciences* 1990;609:87-97.

66. Hoffman RM, Stone SN, Hunt WC, Key CR, Gilliland FD. Effects of misattribution in assigning cause of death on prostate cancer mortality rates. *Annals of Epidemiology* 2003; 13(6):450-4.

67. National Center for Health Statistics. *2003 Revisions of the U.S. Standard Certificate of Live Birth and Death and the Fetal Death Report.* Hyattsville, MD: National Center for Health Statistics; 2005. Available at http://www.cdc.gov/nchs/vital_certs_rev.htm.

68. O'Malley C, Hu KU, West DW. *North American Association of Central Cancer Registries: Race and Ethnicity Identifier Assessment Project.* Springfield, IL: North American Association of Central Cancer Registries; 2001. Available at http://www.naaccr.org/filesystem/pdf/FinalReportofJune82001meeting.pdf.

69. Rosenberg HM, Maurer JD, Sorlie PD, Johnson NJ, MacDorman MF, Hoyert DL, Spitler JF, Scott C. Quality of death rates by race and Hispanic origin: a summary of current research, 1999. *Vital and Health Statistics, Series 2* 1999;(128):1-13.

70. National Center for Health Statistics. *Vital Statistics of the United States: Mortality 1999. The Technical Appendix.* Hyattsville, MD: National Center for Health Statistics; 2004. Available at http://www.cdc.gov/nchs/data/statab/techap99.pdf.

71. Hogan H. The 1990 post-enumeration survey: operations and results. *Journal of the American Statistical Association* 1993;88(423):1047-60.

72. U.S. Cancer Statistics Working Group. *United States Cancer Statistics: 2000 Incidence.* Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention *and* National Cancer Institute; 2003.

73. Reeves T, Bennett C. *The Asian and Pacific Islander Population in the United States: March 2002.* Washington, DC: U.S. Bureau of the Census; 2003. Available at http://www.census.gov/prod/2003pubs/p20-540.pdf.

74. U.S. Bureau of the Census. *The Native Hawaiian and Other Pacific Islander Population. Census 2000 Brief.* Washington, DC: U.S. Bureau of the Census; 2001. Available at http://www.census.gov/prod/2001pubs/c2kbr01-14.pdf.

75. U.S. Bureau of the Census. *The Asian Population. Census 2000 Brief.* Washington, DC: U.S. Bureau of the Census; 2002. Available at http://www.census.gov/prod/2002pubs/c2kbr01-16.pdf.

76. Srinivasan S, Guillermo T. Toward improved health: disaggregating Asian American and Native Hawaiian/Pacific Islander data. *American Journal of Public Health* 2000;90(11): 1731-4.

77. Poe GS, Powell-Griner E, McLaughlin JK, Placek PJ, Thompson GB, Robinson K. Comparability of the death certificate and the 1986 National Mortality Followback Survey. *Vital and Health Statistics, Series 2* 1993;(118): 1-53.

78. Sorlie PD, Rogot E, Johnson NJ. Validity of demographic characteristics on the death certificate. *Epidemiology* 1992;3(2):181-4.

79. U.S. Bureau of the Census. *Overview of Race and Hispanic Origin. Census 2000 Brief.* Washington, DC: U.S. Bureau of the Census; 2001. Available at http://www.census.gov/prod/2001pubs/c2kbr01-1.pdf.

80. U.S. Cancer Statistics Working Group. *United States Cancer Statistics: 2001 Incidence and Mortality.* Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention *and* National Cancer Institute; 2004.

81. U.S. Bureau of the Census. *The Hispanic Population. Census 2000 Brief.* Washington, DC: U.S. Bureau of the Census; 2001. Available at http://www.census.gov/prod/2001pubs/c2kbr01-3.pdf.

82. U.S. Bureau of the Census. *Hispanics and Asian Americans Increasing Faster Than Overall Population.* Washington, DC: U.S. Bureau of the Census; 2005. Available at http://www.census.gov/Press-Release/www/releases/archives/race/001839.html.

83. Ramirez RR, de la Cruz GP. *The Hispanic Population in the United States: March 2002.* Washington, DC: U.S. Bureau of the Census; 2003. Available at http://www.census.gov/prod/2003pubs/p20-545.pdf.

84. Howe HL, Carozza S, O'Malley C, Dolecek TA, Finch JL, Kohler B, West DW, Liu L, Schymura M, Williams M, Abe T, Agovino P, Chen VW, Firth R, Harkins D, Lake A, Roney D, Suarez L. *Cancer in U.S. Hispanics/Latinos, 1995–2000.* Springfield, IL: North American Association of Central Cancer Registries; 2003. Available at http://www.naaccr.org/filesystem/pdf/CIUSHL%202003%20incd.sec0.pdf.

85. NAACCR Expert Panel in Hispanic Identification. *Report of the NAACCR Expert Panel on Hispanic Identification.* Springfield, IL: North American Association of Central Cancer Registries; 2003. Available at http://www.naaccr.org/filesystem/pdf/Hispanic%20Identification%20Report%202003rev%2010-07-03.pdf.

86. NAACCR Latino Research Work Group. *NAACCR Guideline for Enhancing Hispanic/Latino Identification: Revised NAACCR Hispanic/Latino Identification Algorithm [NHIA v2].* Springfield, IL: North American Association of Central Cancer Registries; 2005. Available at http://www.naaccr.org.

87. U.S. Department of the Interior. *Bureau of Indian Affairs.* Washington, DC: U.S. Department of the Interior; 2005. Available at http://www.doi.gov/bureau-indian-affairs.html.

88. Sugarman JR, Holliday M, Ross A, Castorina J, Hui Y. Improving American Indian cancer data in the Washington State Cancer Registry using linkages with the Indian Health Service and tribal records. *Cancer* 1996;78(7 Suppl):1564-8.

89. U.S. Bureau of the Census. *Statistical Abstract of the United States: 2003 (123rd Edition)*. Washington, DC: U.S. Bureau of the Census; 2003. Available at http://www.census.gov/prod/2004pubs/03statab/pop.pdf.

90. Frost F, Taylor V, Fries E. Racial misclassification of Native Americans in a Surveillance, Epidemiology, and End Results cancer registry. *Journal of the National Cancer Institute* 1992;84(12):957-62.

91. Kwong SL, Perkins CL, Snipes KP, Wright WF. Improving American Indian cancer data in the California Cancer Registry by linkage with the Indian Health Service. *Journal of Registry Management* 1998;25(1):17-20.